

# Reproducibility of Joint Swelling Assessment by Sonography in Patients with Long-lasting Rheumatoid Arthritis (SEA-Repro Study Part II)

SANDRINE JOUSSE-JOULIN, MARIA ANTONIETTA d'AGOSTINO, THIERRY MARHADOUR, JEAN DAVID ALBERT, JACQUES BENTIN, ISABELLE CHARY VALCKENAERE, FABIEN ETCHEPARE, PHILIPPE GAUDIN, CHRISTOPHE HUDRY, GÉRARD CHALÈS, LAURENT GRANGE, CÉCILE HACQUARD, DAMIEN LOEUILLE, JÉRÉMIE SELLAM, MAXIME DOUGADOS, and ALAIN SARAUX

**ABSTRACT. Objective.** To evaluate the intraobserver and interobserver reproducibility of B-mode and power Doppler (PD) sonography in patients with active long-standing rheumatoid arthritis (RA) comparatively with clinical data.

**Methods.** In each of 7 patients being considered for a change in their RA treatment regimen, 7 healthcare professionals examined the 28 joints used in the Disease Activity Score 28-joint count (DAS28). Then 7 sonographers examined each of the 7 patients twice, using previously published B-mode and PD grading systems. The clinical reference standard was presence of synovitis according to at least 4/7 examiners. The sonographic reference standard was at least grade 1 (ALG1) or 2 (ALG2) synovitis according to at least 4/7 sonographers. Interobserver reproducibility of sonography was assessed versus the sonographer having the best intraobserver reproducibility. Agreement was measured by Cohen's kappa statistic.

**Results.** Intraobserver and interobserver reproducibility of B-mode and PD used separately was fair to good. Agreement between clinicians and sonographers at all sites using B-mode, PD, and both was 0.46, 0.37, and 0.36, respectively, for grade 1 synovitis; and 0.58, 0.19, and 0.19 for grade 2 synovitis. The number of joints with synovitis was smaller by physical examination (36.7%) than by B-mode with ALG1 (58.6%;  $p < 0.001$ ). The number of joints with synovitis was higher by physical examination than by PD with both ALG1 (17.8%;  $p < 0.0001$ ) and ALG2 (6.6%;  $p < 0.0001$ ).

**Conclusion.** PD findings explain most of the difference between clinical and sonographic joint assessments for synovitis in patients with long-standing RA. (J Rheumatol First Release March 15 2010; doi:10.3899/jrheum.090881)

## Key Indexing Terms:

ULTRASOUND    RHEUMATOID ARTHRITIS    SYNOVITIS    DISEASE ACTIVITY

The development of sonography for evaluating inflammatory joint diseases, most notably rheumatoid arthritis (RA), is of considerable interest. B-mode imaging can be used to visualize synovial hypertrophy, fluid collections, and bony erosions<sup>1,2</sup>; power Doppler imaging (PD) provides information on disease activity<sup>3,4</sup>. Sonography is widely available, noninvasive, and inexpensive. However, few studies have compared sonography findings to clinical findings in

patients with RA. Most European rheumatologists assess RA activity by determining the Disease Activity Score (DAS), a composite index computed from the tender and swollen joint counts determined by physical examination, the erythrocyte sedimentation rate (ESR), and an assessment of general health by the patient using a visual analog scale (VAS). In the short 28-joint count version of the DAS, or DAS28, counts are determined for the following joints:

From the Rheumatology Unit, CHU, and University Hospital La Cavale Blanche, Brest; University Hospital Ambroise Paré, Boulogne Billancourt; Hôpital Sud, Rennes; University Hospital Nancy-Brabois, Vandoeuvre-Lès-Nancy; Cochin University Hospital, Paris; University Hospital Bicêtre, Le Kremlin-Bicêtre; Rheumatology Unit, Les Charmilles Hospital, Arpajon; Rheumatology Unit, GREPI TIMC, Imag UMR CNRS 5525, University Hospital, Hôpital Sud A. Michallon, Grenoble, France; and Rheumatology Unit, University Hospital, Brugmann, Brussels, Belgium.

Supported by Abbott France, Paris, France.

S. Jousse-Joulin, MD; T. Marhadour, MD; A. Saraux, MD, PhD, Rheumatology Unit, CHU and University Hospital La Cavale Blanche; M.A. d'Agostino, MD, PhD; C. Hacquard, MD; Rheumatology Unit, CHU and University Hospital Ambroise Paré; J.D. Albert, MD; G. Chalès, MD,

Rheumatology Unit, CHU and University Hospital, Hôpital Sud; J. Bentin, MD, Rheumatology Unit, University Hospital, Brugmann, Brussels; I. Chary Valckenaere, MD, PhD; D. Loeuille, MD, PhD, Rheumatology Unit, CHU and University Hospital Nancy-Brabois; F. Etchepare, MD, Rheumatology Unit, Les Charmilles Hospital; P. Gaudin, MD, PhD; L. Grange, MD, Rheumatology Unit, GREPI TIMC, Imag UMR CNRS 5525, University Hospital, Hôpital Sud A. Michallon; C. Hudry, MD; M. Dougados, MD, Rheumatology Unit, CHU and Cochin University Hospital; J. Sellam, MD, Rheumatology Unit, CHU, University Hospital Bicêtre.

Address correspondence to Prof. A. Saraux, Rheumatology Unit, La Cavale Blanche Hospital, Brest Teaching Hospital, BP 814, F29609 Brest Cedex, France. E-mail: Alain.Saraux@chu-brest.fr

Accepted for publication December 22, 2009.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2010. All rights reserved.

the metacarpophalangeal joints (MCP), proximal interphalangeal joints (PIP), wrists, elbows, shoulders, and knees. The DAS28 score can be used to guide treatment decisions. Joint swelling is the only presumably objective clinical variable used to compute the DAS28. Nevertheless, classification of a joint as swollen or not swollen may be difficult and partly subjective, so that both intraobserver and interobserver variability occur for the swollen joint count (SJC). Variability in the SJC related to poor reproducibility may lead to fluctuations in the DAS28 score and therefore to inappropriate treatment decisions. Conceivably, reproducibility may improve when joint swelling is determined by sonography instead of physical examination<sup>5</sup>. However, in a study of 15 patients with RA and 3 healthy controls, PD was reliable for assessing inflammatory activity compared to magnetic resonance imaging (MRI), but showed only a weak correlation with clinically determined joint swelling<sup>4</sup>. In a companion study<sup>6</sup>, we found that many joints classified as swollen by physical examination were not active by PD. As a result, the SJC determined by B-mode and PD in combination produced far lower DAS28 scores than the SJC determined by physical examination<sup>6</sup>.

Our main objective was to evaluate the intraobserver and interobserver reproducibility of B-mode and PD used to assess synovitis in patients with active RA of more than 4 years' duration. We also evaluated differences in B-mode and PD findings across joint types and according to the score used to define synovitis. We compared B-mode and PD findings to the results of the physical examination. Finally, we measured the time needed for the sonographic evaluation.

## MATERIALS AND METHODS

We included 7 patients with active RA recruited at the rheumatology department of the Brest Teaching Hospital, Brest, France. They met 1987 revised American College of Rheumatology criteria for RA<sup>7</sup>. There were 5 women and 2 men, with a mean age of 57.1 years (SD 6.8) and a mean disease duration of 22.1 years (SD 13.6). Three patients had rheumatoid nodules. All patients were receiving corticosteroids (mean dosage 8.1 mg/day; SD 3.6) and disease-modifying antirheumatic drugs (methotrexate in 6 patients). Two patients were receiving tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) antagonist therapy. Because of high disease activity, a change in treatment (introduction of a TNF antagonist or switch to another TNF antagonist) was being considered for all 7 patients. Mean tender joint count was 9, mean ESR was 23.8 mm/h, mean pain intensity on a 100-mm VAS was 56.8, and mean SJC was 10.3.

*Conduct of the study.* Participants attended a meeting for the assessments. Each patient (whose identity was concealed by wearing a mask and gown) was examined once by each of 7 clinical healthcare professionals from different cities in France and Belgium<sup>6</sup>, then twice by each of 7 French sonographers.

To simulate the conditions of everyday clinical practice, only 5 min were allowed for the physical joint examination in each patient, no instructions were given before the examinations, and there was no training session. The 28 joints used for the DAS28 were assessed for swelling in each of the 7 patients (for a total of 196 joints), and the findings were scored using a semiquantitative scale (0, no synovitis; 1, synovitis unlikely; 2, synovitis probable; and 3, synovitis present). Joints with scores of 2 or 3 were counted as swollen.

Then 7 experienced sonographers examined the 28 joints in each of the 7 patients on 2 separate occasions. The assessment technique was standardized during a consensus meeting organized just before the assessment session (Table 1). During this consensus meeting, the sonographers received training on the joint assessment and information on the study methodology.

Patients wore masks and gowns to avoid being recognized and were asked not to talk about their clinical symptoms with the sonographers. Each sonographer was timed. Two rounds were organized to evaluate intraobserver and interobserver reliability (image acquisition and reading in real time). An Esaote Technos MPX machine with a 12.5-MHz transducer (Esaote Biomedica, Genoa, Italy) was used. Synovitis was defined as the presence of an intraarticular effusion and/or synovial hypertrophy on B-mode images according to the preliminary Outcome Measures in Rheumatology Clinical Trials (OMERACT) definition of sonographic synovitis<sup>8,9</sup> (Table 1). No quantitative measurement was taken. Synovial blood flow was evaluated by PD in each of the 28 joints. PD measurements were adjusted at the lowest permissible pulse repetition frequency (PRF) to maximize sensitivity, which led to PRF values as low as 750 Hz. Low-wall filters were used. Color gain was set just below the level at which color noise appeared in the underlying bone. Grading systems for both B-mode and PD were based on Szkudlarek's semiquantitative method<sup>10</sup>, but were used on the 28 joints in the DAS28. Synovitis was defined as grade 1 or higher by both B-mode (at least synovial thickening bulging over the line linking the tops of the periarticular bones but without extension along the bone diaphysis) and PD (up to 3 discrete spots or 1 confluent spot plus up to 2 discrete spots), as described in Table 1. Active synovitis was defined as intraarticular synovitis on B-mode images with a signal on PD images.

*Reference standards for presence of synovitis (joint swelling).* The clinical reference standard was a score of 2 or 3 according to at least 4 of the 7 clinical examiners.

Several sonographic reference standards were used. For the assessment of interobserver reproducibility, the reference standard was the result obtained by the sonographer who had the best intraobserver reproducibility. To compare the clinical and sonographic results, the sonographic reference standard was synovitis found by at least 4 (50%) of the 7 sonographers and either grade 1 (ALG1) or higher by B-mode and PD or grade 2 (ALG2) or higher by B-mode and PD. Few joints were ALG3 and therefore we did not evaluate this grade separately.

*Statistical analysis.* Results for quantitative variables are reported as means ( $\pm$  SD) and those for qualitative variables as numbers of positive responses per category (percentages). Patient groups were compared using the chi-square test for qualitative variables and the Mann-Whitney U test for quantitative variables. Values of  $p < 0.05$  were considered statistically significant.

Reproducibility was assessed based on Cohen's kappa coefficient, as follows: excellent, 0.80; good, 0.60–0.79; fair, 0.40–0.59; and poor,  $< 0.40$ . As the number of joints was small, the results by joints are given as an indication, but neither confidence intervals nor comparisons between joints are provided.

Statistical tests were performed using the Statistical Package for the Social Sciences (SPSS 13.0, 2005, SPSS Inc, Chicago, IL, USA).

## RESULTS

*Intraobserver and interobserver reproducibility of the sonographic examination of the 28 joints.* Intraobserver and interobserver reproducibility was assessed using 2 definitions of sonographic synovitis: at least grade 1 (ALG1) and at least grade 2 (ALG2) by B-mode or PD. When B-mode and PD were evaluated separately, intraobserver and interobserver reproducibility was fair to good (Table 2). Sonographer 1 had the best intraobserver reproducibility by B-mode (kappa = 0.75 using ALG1) and sonographer 4 had

Table 1. Sonographic evaluation: scanning technique and scoring of synovitis. Synovitis was defined as at least grade 1 by both B-mode and power Doppler.

Joint, and Type of Sonography	Scanning Technique and Scoring System
Glenohumeral joint	Presence or absence of fluid and/or of hypoechoic synovial hypertrophy in the bicipital groove, transducer transversal and longitudinal to the anterior part of the humerus, shoulder in neutral position. Axillary recess, longitudinal scan.
Elbow	Transverse anterior scan and posterior scan with elbow flexed at 90°: presence or absence of fluid and/or of hypoechoic synovial hypertrophy from the humeral capitellum or the coronoid fossa to the joint capsule.
Wrist	Longitudinal and transversal, from the dorsal aspect with the joint in neutral position, probe located on the 3rd finger: presence or absence of fluid and/or of hypoechoic synovial hypertrophy from the bones to the joint capsule.
5 MCP and 5 PIP	Longitudinal scanning of the dorsal MCP and PIP joints in extension: presence or absence of fluid and/or of hypoechoic synovial hypertrophy from the articular bony margin to the joint capsule.
Knee	Longitudinal and transversal, from the suprapatellar recess, in a supine position, with the joint in 30° of flexion: presence or absence of fluid and/or of hypoechoic synovial hypertrophy in the anteroposterior diameter of the suprapatellar recess.
B-mode	Grade 0 or normal: normal joint (no synovial hypertrophy, no joint effusion) Grade 1 or minimal: minimal synovitis (minimal synovial hypertrophy, with or without minimal joint effusion) Grade 2 or moderate: moderate synovitis (moderate synovial hypertrophy with or without minimal or moderate joint effusion) Grade 3 or severe: severe synovitis (severe synovial hypertrophy, with or without severe joint effusion)
Power Doppler	Grade 0: no vessel in the synovium Grade 1: up to 3 discrete spots or 1 confluent spot + up to 2 discrete spots Grade 2: vessel signals in less than 50% of the surface area of the synovium Grade 3: vessel signals in more than 50% of the surface area of the synovium

MCP: metacarpophalangeal joints; PIP: proximal interphalangeal joints.

Table 2. Intraobserver and interobserver reproducibility of synovitis detection by 7 sonographers using OMERACT criteria and B-mode or power Doppler imaging. Synovitis was defined as at least OMERACT grade 1 (ALG1) or OMERACT grade 2 (ALG2).

Reproducibility	Sonographer	B-mode ALG1	B-mode ALG2	Power Doppler ALG1	Power Doppler ALG2
Intraobserver	1	0.75	0.70	0.48	0.25
	2	0.54	0.61	0.50	0.35
	3	0.53	0.70	0.64	0.44
	4	0.37	0.63	0.52	0.77
	5	0.41	0.59	0.52	0.50
	6	0.67	0.62	0.47	0.44
	7	0.67	0.63	0.66	0.63
Interobserver*	2/ref	0.46	0.61	1/ref 0.37	0.24
	3/ref	0.57	0.68	2/ref 0.26	0.60
	4/ref	0.45	0.66	3/ref 0.55	0.48
	5/ref	0.43	0.55	5/ref 0.47	0.44
	6/ref	0.52	0.58	6/ref 0.39	0.39
	7/ref	0.62	0.65	7/ref 0.49	0.48

\* Interobserver reproducibility was assessed using the sonographer who had the best intraobserver reproducibility as the reference. This was sonographer #1 for B-mode imaging and sonographer #4 for power Doppler imaging. OMERACT: Outcome Measures in Rheumatology Clinical Trials.

the best intraobserver reproducibility by PD (kappa = 0.77 using ALG2). Reproducibility was as good for the sonographic examination as for the physical examination, which had kappa values of 0.31 to 0.77 for intraobserver reproducibility and 0.4 to 0.62 for interobserver reproducibility (the reference being ALG2 or ALG3 according to at least 4/7 clinical examiners).

*Intraobserver reproducibility of the sonographic examination of individual joints.* This assessment was performed by

the sonographers who had the best intraobserver reproducibility for the assessment of 28 joints, that is, sonographer 1 by B-mode and sonographer 4 by PD. The results for B-mode and PD are shown in Table 3. Reproducibility of B-mode assessment was good or excellent at all joint sites except the shoulder using ALG2. Reproducibility of PD assessment was poor to fair using ALG1 but was substantially better using ALG2.

*Agreement between clinicians and sonographers on 28*

**Table 3.** Intraobserver reproducibility of synovitis detection at individual joints by B-mode and power Doppler imaging. This evaluation was done by the sonographers who obtained the best intraobserver reproducibility during the examination of the 28 joints (sonographer #1 for B-mode and sonographer #4 for PD). Synovitis was defined as at least OMERACT grade 1 (ALG1) or OMERACT grade 2 (ALG2).

Joints		Sonographer #1 B-mode ALG1			Sonographer #4 Power Doppler ALG1			Sonographer #1 B-mode ALG2			Sonographer #4 Power Doppler ALG2		
		No	Yes	$\kappa$	No	Yes	$\kappa$	No	Yes	$\kappa$	No	Yes	$\kappa$
Elbow	no	10	0	1	14	0	NA	10	0	1	14	0	NA
	yes	0	4		1	0		0	4		0		
Shoulder	no	10	0	0.81	14	0	NA	11	1	-0.07	14	0	NA
	yes	1	3		0	0		1	1		0	0	
Knee	no	5	2	0.60	14	0	NA	9	1	NA	14	0	NA
	yes	0	3		0	0		0	0		0		
PIP	no	19	3	0.75	60	5	0.27	26	1	0.71	67	2	0.48
	yes	2	17		5	2		4	10		0	1	
MCP	no	18	4	0.67	38	13	0.5	21	6	0.63	48	6	0.77
	yes	6	42		3	16		6	37		1	15	
Wrist	no	5	1	0.70	5	3	0.44	6	0	0.58	9	0	0.83
	yes	1	7		1	5		3	5		1	4	

NA: not applicable; PIP: proximal interphalangeal joints; MCP: metacarpophalangeal joints; OMERACT: Outcome Measures in Rheumatology Clinical Trials.

*joints.* Table 4 shows the agreement between joint swelling found by at least 4/7 clinical examiners and synovitis found by at least 4/7 sonographers, using at least ALG1 or at least ALG2 by B-mode or PD on the 28 joints. Agreement between clinical examiners and sonographers was better using ALG2 than using ALG1 by B-mode, while the opposite was true by PD: with B-mode, PD, and both, agreement between synovitis by physical examination and synovitis by sonography at all sites was 0.46, 0.37, and 0.36, respectively, using ALG1; and 0.58, 0.19, and 0.19, respectively, using ALG2. Finally, the number of joints with synovitis was smaller by physical examination than by B-mode when ALG1 was used [72/196 (36.7%; 95% CI 30.1–43.9) for clinicians vs 115/196 (58.6%; 95% CI 51.4–65.5) for sonographers;  $p < 0.001$ ], while the opposite was true when ALG2 was used [65/196 (33.2%; 95% CI 26.7–40.2);  $p = 0.46$ , nonsignificant]. The number of joints with synovitis was higher by physical examination than by PD using both

ALG1 [35/196 (17.8%; 95% CI 12.9–24);  $p < 0.0001$ ] and ALG2 [13/196 (6.6%; 95% CI 3.7–11.3);  $p < 0.0001$ ].

*Agreement between clinicians and sonographers for individual joints.* Table 5 shows the agreement between joint swelling found by at least 4/7 clinical examiners and synovitis found by at least 4/7 sonographers, using at least ALG1 or at least ALG2 by B-mode or PD at each individual joint. With B-mode, agreement was low at the knee (–0.14 and –0.07 using ALG1 and ALG2, respectively) and fair at other sites (0.32 to 0.59). With PD, agreement was low at the proximal interphalangeal joints (0.04 using ALG1) and usually better with ALG1 than with ALG2.

*Mean duration of the sonographic evaluation of each joint group.* Examination of each joint group by both B-mode and PD required  $125 \pm 74$  seconds overall. The mean ranged from  $67 \pm 17$  seconds for the knees to  $255 \pm 78$  seconds for the MCP joints (Table 6).

**Table 4.** Agreement between presence of clinical swelling according to at least 4 of 7 clinical examiners and presence of synovitis according to at least 4 of 7 sonographers for all 28 joints. Synovitis by sonography was defined as at least OMERACT grade 1 (ALG1) or OMERACT grade 2 (ALG2) by B-mode imaging or by power Doppler imaging.

Clinical Synovitis	B-mode		Power Doppler		B-mode and Power Doppler	
	No	Yes	No	Yes	No	Yes
At least ALG1						
No	75	49	116	7	116	7
Yes	6	66	44	28	45	27
$\kappa$	0.46		0.37		0.36	
At least ALG2						
No	109	15	123	1	123	1
Yes	22	50	60	12	60	12
$\kappa$	0.58		0.19		0.19	

OMERACT: Outcome Measures in Rheumatology Clinical Trials.

Table 5. Agreement between presence of clinical swelling according to at least 4 of 7 clinical examiners and presence of synovitis according to at least 4 of 7 sonographers for each individual joint or joint group. Synovitis by sonography was defined as at least OMERACT grade 1 (ALG1) or as at least OMERACT grade 2 (ALG2) by B-mode imaging or by power Doppler imaging.

Clinical Synovitis		ALG1 Synovitis B-mode			ALG1 Synovitis Power Doppler			ALG2 Synovitis B-mode			ALG2 Synovitis Power Doppler		
		No	Yes	$\kappa$	No	Yes	$\kappa$	No	Yes	$\kappa$	No	Yes	$\kappa$
Elbow	no	10	3	0.32	13	0	NA	11	2	0.44	13	0	NA
	yes	0	1		1	0		0	1		0		
Shoulder	no	11	3	NA	14	0	NA	13	1	NA	14	0	NA
	yes	0	0		0	0		0	0		0		
Knee	no	3	10	-0.14	13	0	NA	12	1	-0.07	13	0	NA
	yes	1	0		1	0		1	0				
PIP	no	30	15	0.52	42	2	0.04	43	2	0.59	45	0	NA
	yes	2	23		23	2		10	15		25	0	
MCP	no	18	15	0.50	30	3	0.41	25	8	0.54	33	0	0.18
	yes	2	35		18	19		8	29		30	7	
Wrist	no	3	3	0.39	4	2	0.55	5	1	0.44	5	1	0.44
	yes	1	7		1	7		3	5		3	5	

NA: not applicable; PIP: proximal interphalangeal joints; MCP: metacarpophalangeal joint; OMERACT: Outcome Measures in Rheumatology Clinical Trials.

Table 6. Mean duration of sonography (s) by joint group.

Joint	Mean (s)	SD
Elbow	82	27
Shoulder	115	26
Knee	67	17
PIP	144	36
MCP	255	78
Wrist	90	19
Total	125	74

PIP: proximal interphalangeal joints; MCP: metacarpophalangeal joints.

## DISCUSSION

Several studies have established that sonography can detect synovial membrane thickening, joint effusions, and bone erosions in the MCP joints of patients with RA<sup>3</sup>. Subsequent studies established that the PD signal reflecting blood flow in the synovial membrane correlated with disease activity<sup>1,2</sup>. Sonography has been evaluated comparatively to MRI and histology<sup>4</sup>.

Sonography is widely viewed as heavily operator-dependent. Few studies of intraobserver reproducibility evaluating both acquisition and reading have been published. In a systematic review of the metric properties of sonography performed by the OMERACT Ultrasound Group<sup>11</sup>, a single study of intraobserver reproducibility was identified<sup>12</sup>. More data exist on intraobserver reading reproducibility using pictures or video clips<sup>12-20</sup>. Early studies of sonogram reproducibility in patients with RA tested only 2 sonographers<sup>5,10,21,22</sup>.

We investigated the reproducibility of both B-mode and PD performed by 7 sonographers in 7 patients with RA. Reproducibility was assessed for synovitis defined as a semiquantitative score of 1 (ALG1) or 2 (ALG2).

Intraobserver reproducibility for ALG1 by B-mode was poor for 1 sonographer, fair for 3 sonographers, and good for 3 sonographers. Using ALG2 by B-mode, intraobserver reproducibility was good for 6 sonographers. Reproducibility using ALG1 or ALG2 by PD was good for only 2 sonographers.

Several studies of interobserver acquisition reproducibility have been performed in patients with RA<sup>4,5,12,14,16,19,21,23-28</sup>. In 1 study, 14 sonographers examined 4 patients with inflammatory joint diseases (1 case each of RA, gout, remitting seronegative symmetrical synovitis with pitting edema, and reactive arthritis)<sup>19</sup>. Kappa values were low for the detection of synovitis and effusion. In contrast, a study of 23 sonographers and 24 patients (including 3 with RA) found high kappa values for effusion/synovitis at the wrist/hand (0.73) and ankle/foot (0.69)<sup>12</sup>.

Intraobserver and interobserver reproducibility of PD was tested using video clips of healthy joints and of joints from patients with monoarthritis or polyarthritis<sup>24</sup>. The clips were sent to 17 sonographers who were asked whether a Doppler signal was present and how they scored the intensity of the signal on a scale from 0 to 3. Intraobserver reproducibility of signal detection and scoring was good (kappa = 0.82; 0.58–0.96) and interobserver reproducibility was moderate (kappa = 0.66). The most difficult joints to assess were the knee, MCP, wrists, and elbows.

In our study, the reproducibility of both B-mode and PD was low, although the sonographers were experienced in the technique. One explanation may be that the sonographers did not use their usual sonography machine. Another possibility is that joint palpation during the 14 clinical evaluations followed by pressure from the probe during 14 sonographic evaluations may modify the results. Interestingly, when the same sonographers evaluated their intraobserver

reliability with other patients using their own sonography machines to examine the joints twice at a 2-day interval, intraobserver reliability ranged from 0.61 to 0.97<sup>29</sup>, compared to 0.37 to 0.75 in our study. Thus, reproducibility in our study may have been lower than in routine practice.

We evaluated interobserver reproducibility using the findings by the sonographer who had the best intraobserver reproducibility as the reference standard. For diagnosing ALG1 synovitis on B-mode images, reproducibility was fair for 5 of 6 sonographers and good for 1 sonographer. For ALG2 synovitis, reproducibility was good for 4 of 6 sonographers and fair for 2. For diagnosing ALG1 inflammation by PD, agreement was poor for 3 sonographers and fair for 3 sonographers. For ALG2 inflammation by PD, agreement was good for 1 sonographer, fair for 3 sonographers, and poor for 2 sonographers. Thus, reproducibility was better with ALG2 than with ALG1 by both B-mode and PD.

In contrast, we found no marked differences in reproducibility across joints. Reproducibility of the B-mode assessment was good or excellent at all joints except the shoulder when ALG2 was used to define synovitis. By PD, reproducibility was better using ALG2 than ALG1, but with both ALG2 and ALG1 reproducibility was better at the MCP and wrist than at the PIP joints.

We also evaluated agreement between the clinical examiners and sonographers regarding the presence of synovitis. Agreement was best between clinical ALG2 and BM ALG2 at the PIP joints and MCP joints.

Agreement was poor at the elbows and wrists in our study and in earlier work<sup>10,30,31</sup>. In a study of agreement between 2 sonographers examining patients with lupus, the kappa value for detecting wrist synovitis was 0.73<sup>30</sup>. In a study of 50 patients with RA and 20 healthy controls, agreement was only fair between elbow swelling detected clinically by a rheumatologist and joint effusion detected by a sonographer<sup>32</sup>. Changes induced by intraarticular glucocorticoid injections in 20 patients with chronic synovitis of the hands, feet, or wrists (including 11 with RA) were assessed by 1 rheumatologist and 2 sonographers<sup>22</sup>. Kappa values were 0.86 for detection of effusion/synovitis and 0.95 for detection of a PD signal<sup>22</sup>.

At the shoulders, synovitis was detected more often by B-mode than by physical examination. The shoulder is known to be difficult to examine<sup>33-35</sup>. At the knees, synovitis was detected more often by B-mode with ALG1 than by physical examination. Small effusions may be more difficult to detect at the knee by physical examination than by sonography<sup>36,37</sup>.

Agreement was slightly lower with ALG1 than with ALG2 by B-mode, while the opposite was true by PD. Agreement between clinical swelling and presence of a PD signal was best at the wrists. None of the patients had PD signals at the elbows, shoulders, or knees.

In a study of sonography to evaluate the fingers and toes

of patients with RA, an experienced radiologist and a rheumatologist with limited training in sonography obtained good agreement for detecting synovitis<sup>10</sup>. Kappa values for the semiquantitative assessment of effusions, synovitis, PD, and erosions were 0.79/0.48, 0.86/0.63, 0.87/0.55, and 0.91/0.68, respectively<sup>10</sup>.

In our study, the number of joints with synovitis (SJC) was higher by B-mode with ALG1 than by physical examination at the elbows, shoulders, knees, PIP joints, MCP joints, and wrists. That sonography can detect subclinical synovitis has been established<sup>31,32</sup>. In contrast, the clinical SJC was higher than the number of joints generating a PD signal. It is important to note that our patients had long-standing disease with joint erosions. In joints with advanced disease, the presence and the activity of synovitis may be more difficult to detect than at earlier stages. Presence of a PD signal was perhaps a more reliable sign of activity than clinical joint swelling.

The mean time needed for the sonographic assessment in our study was  $125 \pm 74$  seconds per joint group overall and ranged from  $67 \pm 17$  seconds for the knees to  $255 \pm 78$  seconds for the MCP joints. Studies are needed to determine whether evaluating selected joint groups or using only B-mode or PD can decrease the examination time while providing useful and reproducible information.

Intraobserver reproducibility is as good for sonographic evaluation of synovitis as for clinical detection of joint swelling. With both B-mode and PD, and using either ALG1 or ALG2, the number of joints with synovitis was smaller by sonography than by physical examination. In contrast, when only B-mode was used, with ALG1, the number was larger by sonography. Agreement between physical examination and B-mode ALG2 (kappa = 0.58) was as good as reproducibility among clinical examiners (kappa = 0.40 to 0.62) and slightly lower than reproducibility among sonographers using B-mode (kappa = 0.55 to 0.68). Most of the differences between the results of the physical and sonographic evaluations were related to the PD findings. Studies are therefore needed to evaluate the usefulness of PD using ALG1 only for the evaluation of patients with RA in everyday practice.

Most clinical trials have used clinician SJC for patient evaluation. We recently demonstrated that sonographic evaluation of synovitis was at least as relevant an outcome measure as physical examination<sup>29</sup>. Here, we found that B-mode ALG2 gives results quite similar to clinical evaluation, while B-mode ALG1 detects a higher number, and PD a lower number, of joints with synovitis. Before accepting sonography for RA evaluation, we have to determine which criterion is best for identifying clinically relevant synovitis. If we use B-mode plus PD, the SJC will be lower with sonography than with clinician evaluation in long-lasting RA, which may significantly reduce the DAS28, as shown in Marhadour, *et al*<sup>6</sup>.

## REFERENCES

1. Newman JS, Laing TJ, McCarthy CJ, Adler RS. Power Doppler sonography of synovitis: assessment of therapeutic response — preliminary observations. *Radiology* 1996;198:582-4.
2. Hau M, Schultz H, Tony HP, Keberle M, Jahns R, Haerten R, et al. Evaluation of pannus and vascularization of the metacarpophalangeal and proximal interphalangeal joints in rheumatoid arthritis by high-resolution ultrasound (multidimensional linear array). *Arthritis Rheum* 1999;42:2303-8.
3. Grassi W, Tittarelli E, Pirani O, Avaltroni D, Cervini C. Ultrasound examination of metacarpophalangeal joints in rheumatoid arthritis. *Scand J Rheumatol* 1993;22:243-7.
4. Szkudlarek M, Court-Payen M, Strandberg C, Klarlund M, Klausen T, Ostergaard M. Power Doppler ultrasonography for assessment of synovitis in the metacarpophalangeal joints of patients with rheumatoid arthritis: a comparison with dynamic magnetic resonance imaging. *Arthritis Rheum* 2001;44:2018-23.
5. Karim Z, Wakefield RJ, Quinn M, Conaghan PG, Brown AK, Veale DJ, et al. Validation and reproducibility of ultrasonography in the detection of synovitis in the knee: a comparison with arthroscopy and clinical examination. *Arthritis Rheum* 2004;50:387-94.
6. Marhadour T, Jousse-Joulin S, Chalès G, Grange L, Hacquard C, Loeuille D, et al. Reproducibility of joint swelling assessments in long-lasting rheumatoid arthritis: Influence on Disease Activity Score-28 values (SEA-Repro Study Part I). *J Rheumatol* 2010;37:[in press].
7. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315-24.
8. Wakefield RJ, Balint PV, Szkudlarek M, Filippucci E, Backhaus M, d'Agostino MA, et al, OMERACT 7 Special Interest Group. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. *J Rheumatol* 2005;32:2485-7.
9. Wakefield RJ, d'Agostino MA, Iagnocco A, Filippucci E, Backhaus M, Scheel AK, et al, OMERACT Ultrasound Group. The OMERACT ultrasound group: status of current activities and research directions. *J Rheumatol* 2007;34:848-51.
10. Szkudlarek M, Court-Payen M, Jacobsen S, Klarlund M, Thomsen HS, Østergaard M. Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis. *Arthritis Rheum* 2003;48:955-62.
11. Joshua F, Lassere M, Bruyn GA, Szkudlarek M, Naredo E, Schmidt WA, et al. Summary findings of a systematic review of the ultrasound assessment of synovitis. *J Rheumatol* 2007;34:839-47.
12. Naredo E, Möller I, Moragues C, de Agustín JJ, Scheel AK, Grassi W, et al. Interobserver reliability in musculoskeletal ultrasonography: results from a “teach the teachers” rheumatologist course. *Ann Rheum Dis* 2006;65:14-9.
13. Backhaus M, Kamradt T, Sandrock D, Loreck D, Fritz J, Wolf KJ, et al. Arthritis of the finger joints: a comprehensive approach comparing conventional radiography, scintigraphy, ultrasound, and contrast-enhanced magnetic resonance imaging. *Arthritis Rheum* 1999;42:1232-45.
14. Wakefield RJ, Gibbon WW, Conaghan PG, O'Connor P, McGonagle D, Pease C, et al. The value of sonography in the detection of bone erosions in patients with rheumatoid arthritis: a comparison with conventional radiography. *Arthritis Rheum* 2000;43:2762-70.
15. Qvistgaard E, Røgind H, Torp-Pedersen S, Terslev L, Danneskiold-Samsøe B, Bliddal H. Quantitative ultrasonography in rheumatoid arthritis: evaluation of inflammation by Doppler technique. *Ann Rheum Dis* 2001;60:690-3.
16. Ribbens C, André B, Marcelis S, Kaye O, Mathy L, Bonnet V, et al. Rheumatoid hand joint synovitis: gray-scale and power Doppler US quantifications following anti-tumor necrosis factor-alpha treatment: pilot study. *Radiology* 2003;229:562-9.
17. Varsamidis K, Varsamidou E, Tjetjis V, Mavropoulos G. Doppler sonography in assessing disease activity in rheumatoid arthritis. *Ultrasound Med Biol* 2005;31:739-43.
18. Frediani B, Falsetti P, Storri L, Bisogno S, Baldi F, Campanella V, et al. Evidence for synovitis in active polymyalgia rheumatica: sonographic study in a large series of patients. *J Rheumatol* 2002;29:123-30.
19. Scheel AK, Schmidt WA, Hermann KG, Bruyn GA, d'Agostino MA, Grassi W, et al. Interobserver reliability of rheumatologists performing musculoskeletal ultrasonography: results from a EULAR “train the trainers” course. *Ann Rheum Dis* 2005;64:1043-9.
20. Iagnocco A, Cerioni A, Coari G, Ossandon A, Masciangelo R, Valesini G. Intra-articular methotrexate in the treatment of rheumatoid arthritis and psoriatic arthritis: a clinical and sonographic study. *Clin Rheumatol* 2006;25:159-63.
21. D'Agostino M, Maillefert J, Said-Nahal R, Breban M, Ravaud P, Dougados M. Detection of small joint synovitis by ultrasonography: the learning curve of rheumatologists. *Ann Rheum Dis* 2004;63:1284-7.
22. Filippucci E, Farina A, Carotti M, Salaffi F, Grassi W. Grey scale and power Doppler sonographic changes induced by intra-articular steroid injection treatment. *Ann Rheum Dis* 2004;63:740-3.
23. Hoving JL, Buchbinder R, Hall S, Lawler G, Coombs P, McNealy S, et al. A comparison of magnetic resonance imaging, sonography, and radiography of the hand in patients with early rheumatoid arthritis. *J Rheumatol* 2004;31:663-75.
24. Koski JM, Saarakkala S, Helle M, Hakulinen U, Heikkinen JO, Hermunen H, et al. Assessing the intra- and inter-reader reliability of dynamic ultrasound images in power Doppler ultrasonography. *Ann Rheum Dis* 2006;65:1658-60.
25. Naredo E, Rodríguez M, Campos C, Rodríguez-Heredia JM, Medina JA, Giner E, et al. Validity, reproducibility, and responsiveness of a twelve-joint simplified power doppler ultrasonographic assessment of joint inflammation in rheumatoid arthritis. *Arthritis Rheum* 2008;59:515-22.
26. Bruyn GA, Naredo E, Möller I, Moragues C, Garrido J, de Bock GH, et al. Reliability of ultrasonography in detecting shoulder disease in patients with rheumatoid arthritis. *Ann Rheum Dis* 2009;68:357-61.
27. Fiocco U, Ferro F, Cozzi L, Vezzù M, Sfriso P, Checchetto C, et al. Contrast medium in power Doppler ultrasound for assessment of synovial vascularity: comparison with arthroscopy. *J Rheumatol* 2003;30:2170-6.
28. Fiocco U, Ferro F, Vezzù M, Cozzi L, Checchetto C, Sfriso P, et al. Rheumatoid and psoriatic knee synovitis: clinical, grey scale, and power Doppler ultrasound assessment of the response to etanercept. *Ann Rheum Dis* 2005;64:899-905.
29. Dougados M, Jousse-Joulin S, Mistretta F, d'Agostino MA, Backhaus M, Bentin J, et al. Evaluation of several ultrasonography scoring systems of synovitis and comparison to clinical examination: Results from a prospective multi-center study of rheumatoid arthritis. *Ann Rheum Dis* 2009 Sep 9. (E-pub ahead of print)
30. Iagnocco A, Ossandon A, Coari G, Conti F, Priori R, Alessandri C, et al. Wrist joint involvement in systemic lupus erythematosus. An ultrasonographic study. *Clin Exp Rheumatol* 2004;22:621-4.
31. Luukkainen R, Sanila MT, Saltyshv M, Huhtala H, Koski JM. Relationship between clinically detected joint swelling and effusion diagnosed by ultrasonography in elbow joints in patients with rheumatoid arthritis. *Clin Rheumatol* 2005;24:228-31.
32. Karim Z, Wakefield RJ, Conaghan PG, Lawson CA, Goh E, Quinn MA, et al. The impact of ultrasonography on diagnosis and

- management of patients with musculoskeletal conditions. *Arthritis Rheum* 2001;44:2932-3.
33. Wakefield RJ, Green MJ, Marzo-Ortega H, Conaghan PG, Gibbon WW, McGonagle D, et al. Should oligoarthritis be reclassified? Ultrasound reveals a high prevalence of subclinical disease. *Ann Rheum Dis* 2004;63:382-5.
  34. Luukkainen R, Sanila MT, Luukkainen P. Poor relationship between joint swelling detected on physical examination and effusion diagnosed by ultrasonography in glenohumeral joints in patients with rheumatoid arthritis. *Clin Rheumatol* 2007;26:865-7.
  35. Naredo E, Aguado P, De Miguel E, Uson J, Mayordomo L, Gijon-Baños J, et al. Painful shoulder: comparison of physical examination and ultrasonographic findings. *Ann Rheum Dis* 2002;61:132-6.
  36. Hauzeur JP, Mathy L, De Maertelaer V. Comparison between clinical evaluation and ultrasonography in detecting hydrarthrosis of the knee. *J Rheumatol* 1999;26:2681-3.
  37. Kane D, Balint PV, Sturrock RD. Ultrasonography is superior to clinical examination in the detection and localization of knee joint effusion in rheumatoid arthritis. *J Rheumatol* 2003;30:966-71.